PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12N 9/50, 15/51, 15/62, C07K 14/18 | A1

(11) International Publication Number: WO 98/12308

(43) International Publication Date: 26 March 1998 (26.03.98)

(21) International Application Number: PCT/IT97/00228

(22) International Filing Date: 17 September 1997 (17.09.97)

(30) Priority Data:
RM96A000632 17 September 1996 (17.09.96) IT

(71) Applicant (*for all designated States except US*): ISTITUTO DI RICERCHE DI BIOLOGIA MOLECOLARE P. ANGELETTI S.P.A [IT/IT]; Via Pontina Km. 30,600, I-00040 Pomezia RM (IT).

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): DE FRANCESCO, Raffaele [IT/IT]; Via Appia Nuova, Km.21 N.133, I-00040 Marino (IT). TRAMONTANO, Anna [IT/IT]; Via E. Mattei, 19, I-00045 Genzano (IT). TOMEI, Licia [IT/IT]; Via C.E. Gadda, 173, I-00143 Rome (IT). NARDI, Maria, Chiara [IT/IT]; Via delle Pesche, 4, I-00040 Pomezia RM (IT). STEINKÜHLER, Christian [DE/IT]; Via Vitaschi, 17, I-00152 Rome (IT). URBANI, Andrea [IT/IT]; Via Piave, 71, I-00187 Rome (IT). VITALE, Rosa, Letizia [IT/IT]; Via U. La Malfa 15, Scala B, I-00040 Pomezia RM (IT). COLLOCA, Stefano [IT/IT]; Via Peretti, 22, I-00153 Rome (IT).

(74) Agents: DI CERBO, Mario et al.; Società Italiana Brevetti S.p.A., Piazza di Pietra, 39, I-00186 Roma (IT).

(81) Designated States: AU, CA, CN, IL, JP, KR, MX, US, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: SOLUBLE POLYPEPTIDES WITH ACTIVITY OF THE NS3 SERINE PROTEASE OF HEPATITIS C VIRUS, AND PROCESS FOR THEIR PREPARATION AND ISOLATION

(57) Abstract

The present invention relates to serine protease NS3 of hepatitis C virus, and in particular to the observation that the NS3 serine protease domain, in its native conformation, binds a Zn^{2+} ion and that bivalent metallic ions are necessary to the structural integrity of the protein and to the activity of the enzyme. The present invention further relates to recombinant polypeptides which comprise sequences of the NS3 protease and are characterised by a tail of at least three lysines at their C-terminal ends, to increase its solubility. A further subject of the present invention is a new process which allows the expression of said polypeptides, as metalloproteins, with the proteolytic activity of the HCV NS3 protease, in a soluble form and in a quantity sufficient to allow research to identify inhibitors and to determine the three-dimensional structure of the NS3 protease. Figure 4 shows the effects of the zinc ion on the production of the HCV NS3 protease as a soluble protein in E. Coli in a minimum culture medium.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

SOLUBLE POLYPEPTIDES WITH ACTIVITY OF THE NS3 SERINE PROTEASE OF HEPATITIS C VIRUS, AND PROCESS FOR THEIR PREPARATION AND ISOLATION

DESCRIPTION

The hepatitis C virus (HCV) is the main etiologic agent of non-A, non-B hepatitis (NANB). It is estimated that HCV causes at least 90% of post-transfusional NANB viral hepatitis and 50% of sporadic NANB hepatitis. Although great progress has been made in the selection of blood donors and in the immunological characterisation of blood used for transfusions, there is still a high level of acute HCV infection among those receiving blood transfusions, resulting in one million or more infections every year throughout the world. Approximately 50% of HCV infected individuals develop cirrhosis of the liver within a period that can range from 5 to 40 years, and recent clinical studies suggest that there is a correlation between chronic HCV infection and the development of hepatocellular carcinoma.

HCV is an enveloped virus containing an RNA positive genome of approximately 9.4 kb. This virus is a member of the *Flaviviridae* family, the other members of which are the pestiviruses and flaviviruses.

The RNA genome of HCV has recently been sequenced. Comparison of sequences from the HCV genomes isolated in various parts of the world has shown that these sequences can be extremely heterogeneous. Most of the HCV genome is occupied by an open reading frame (ORF) that can vary between 9030 and 9099 nucleotides. This ORF codes for a single viral polyprotein, the length of which can obviously vary from 3010 to 3033 amino acids. During the virus infection cycle, the polyprotein is proteolytically processed into the individual gene products necessary for replication of the virus.

The genes coding for HCV structural protein are located at the 5' end of the ORF, whereas the region coding for the non-structural proteins occupies the rest

of the ORF. The structural proteins consist of: C (core, 21 kDa), E1 (envelope, gp37) and E2 (NS1, gp61). C is a non-glycosilate protein of 21 kDa, which probably forms the viral nucleocapsid. The protein E1 is a glycoprotein of approximately 37 kDa and is believed to be a structural protein of the outer viral envelope. E2, another membrane glycoprotein of 61 kDa, is probably a second structural protein of the outer envelope of the virus.

The non-structural region starts with NS2 (p24), a hydrophobic protein of 24 kDa whose function is not known. NS3, a protein of 68 kDa which follows NS2 in the polyprotein, has two functional domains: a serine protease domain in the first 180 amino-terminal amino acids and an RNA-dependent ATPase domain in the carboxy-terminal part. The gene region corresponding to NS4 codes for NS4A (p6), a membrane protein of 54 amino acids, and NS4B (p26). The gene corresponding to NS5 codes for two proteins, NS5A (p56) and NS5B (p65), of 56 and 65 kDa, respectively. Recently it has been shown that the NS5B region has an RNA dependent RNA-polymerase activity (1).

Various molecular biological studies indicate that the signal peptidase, a protease associated with the endoplasmic reticulum of the host cell, is responsible for proteolytic processing in the non-structural region, that is to say the sites C/E1, E1/E2 and E2/NS2 (2). A first protease activity of HCV is responsible for the cleavage between NS2 and NS3. This activity is contained in a region comprising both a part of NS2 and the part of NS3 containing the serine protease domain, but does not use the same catalytic mechanism (3). On the contrary, the serine protease contained in the 180 amino acids at the amino-terminal of NS3 is responsible for cleavage at the junctions between NS3 and NS4A, between NS4A and NS4B, between NS4B and NS5A, and between NS5A and NS5B (4-8). In particular it has been found that the cleavage produced by this serine protease leaves a residue of

cysteine or threonine on the amino-terminal side (position P1) and a residue of alanine or serine on the carboxy-terminal side (position P1') of the substrate (6, 9). Recently it has been shown that NS4A binds the N-terminal end of NS3 with its central hydrophobic portion, thereby increasing the proteolytic activity of NS3 in all the cleavage sites on the polyprotein (10-12).

Inhibition of the protease activity would therefore stop the proteolytic processing of the non-structural portion of the HCV polyprotein and, as a consequence, would prevent virus replication in infected cells. This sequence of events has been verified in a flavivirus, homologous of the hepatitis C virus, which infects cells in culture.

In this case it has been possible to show that genetic manipulation, producing a protease that is no longer capable of exerting its catalytic activity, abolishes the ability of the virus to replicate (13). Furthermore it has been widely demonstrated, both in *vitro* and in clinical studies, that compounds capable of interfering with the activity of the HIV protease are capable of inhibiting the replication of this virus (14).

Finally there is evidence of the fact that the NS5 region of HCV, which as we have mentioned above has an RNA dependent RNA-polymerase activity, does not display this function except after processing by the NS3 protease.

Therefore a substance capable of interfering with the proteolytic activity associated with the NS3 protein, could be a new therapeutic agent. From this point of view detailed knowledge of the three-dimensional structure of the protease takes on a great deal of importance, as it would allow both a greater understanding of the biological phenomena in which it is involved, and the analysis, study and design of inhibitor molecules capable of interfering with the protease activity, thus paving the way for the development of pharmaceutical

compositions suitable for treatment of hepatitis C. Nevertheless, determination of the structure both using NMR methods and X-ray crystallography, requires large amounts of soluble protein, and at the present time it is not possible to meet this request. In fact, although the simplest and most economical manner of obtaining large amounts of the desired polypeptide is expression of the corresponding gene in bacteria, and although there is a widespread availability of numerous eucaryotic promoters and methods for maximising the expression of heterologous genes in E. Coli, nevertheless an efficient production of the polypeptide in question, although necessary, might not be sufficient. Many recombinant proteins do not fold the polypeptidic chain correctly when they are expressed in E. Coli. The result is the synthesis of polypeptides which are either degraded in the host cell, or are accumulated in an insoluble form in the so called inclusion bodies (15). Furthermore, in the case of extremely hydrophobic proteins, proteins of viral origin or proteins that are toxic for the bacterial cell (as is the case for certain proteases of viral origin) there are insurmountable difficulties in producing them in a native, soluble form.

In the case of the NS3 serine protease of the hepatitis C virus, due to the conditions in which the protein is normally produced, it has not been possible to date to obtain in E. coli a native type, soluble protease in amounts sufficient to enable the study of the structural nature of this protein, which requires solutions containing a high millimolar concentration of the protein.

It has now unexpectedly been found that these important limitations can be overcome by using the method according to the present invention. As will be seen from the following, this method is based on the unexpected discovery that the NS3 serine protease domain, in its native conformation, binds a Zn^{2+} ion.

Because, as mentioned above, the structure of the HCV NS3 protease is not yet known, a structural model of the protein was prepared, to be used as a guide during experiments. However, the similarity of the NS3 protease to other serine proteases of known structure is extremely low (less than 15%), which does not allow good alignment between sequences and as a result does not allow construction of a three-dimensional model based solely on homology. For this reason, the available serine protease structures were used to build a multiple alignment of the structurally conserved regions and to draw up in this way a profile with which the sequence of the NS3 protease could subsequently be aligned. In this way it was possible to build an approximate three-dimensional model of the HCV NS3 protease (9, 16).

Recently, three new viruses responsible for human hepatitis have been discovered (17). These new viruses, known as GBV-A, GBV-B and GBV-C, show a polyprotein organisation in common with that of HCV (18, 19). From alignment of the region corresponding to NS3 in these three new viruses with that of various HCV serotypes, several preserved amino acids were identified. These residues comprise: the amino acids in the active site, some glycines and prolines (probably involved in stabilising the structure of the protein) and three cysteines and one histidine (figure 1). In the model suggested by us for the NS3 protease these last four residues are found in a region of the molecule opposite the active site, in a close spatial relationship, and their relative position is such that it forms a binding site for a divalent metallic ion, such as for example the ion Zn^{2+} (figure 2).

This observation was subsequently confirmed experimentally. In fact, as will be illustrated in greater detail in the examples, the HCV NS3 protease actually has a metal content equivalent to one mole of zinc to each mole of protein, and as is the case in other

proteins the zinc is necessary to enable the protein to take on its native structure and become catalytically active (20, 21).

The fact that the NS3 protease has a binding site for a metal ion and that this binding site is so well preserved, even in viruses that are not phylogenetically close, opens the way to the study of antiviral therapeutic agents whose target site is this very region of the protein. In fact, in the case of another viral protein that binds Zn^{2+} ions, that is to say the HIV virus nucleocapsid, it has been possible to identify compounds that interfere selectively with the bond between the protein and the Zn^{2+} ions (22, 23) and it has also been seen that these compounds interfere with the viral infection of cells grown in culture medium.

An object of the present invention is therefore to provide a method for high-yield expression, in a native form, that is to say as a protein containing a bivalent metallic ion, and in a highly soluble form of the HCV NS3 protease using heterologous expression systems, such as E. coli cells transformed using suitable genetic constructs and cultivated in a medium enriched with salts containing divalent metal ions.

A further object of the present invention is to provide a general method allowing preparation and isolation in a native, pure and highly soluble form, of large amounts of polypeptides containing Zn^{2+}, Co^{2+} or Cd^{2+}, with the protease activity of HCV NS3.

Furthermore, an additional object of the present invention is to provide a method that allows preparation and isolation in a native, pure and highly soluble form of large amounts of polypeptides with the protease activity of HCV NS3, which are at the same time marked using stable heavy isotopes such as 13_C or 15_N, as required for experiments to determine the three-dimensional structure of the protein using NMR.

Finally, the present invention provides new genetic constructs for the expression, in E. coli cells, of modified polypeptides with the protease activity of HCV NS3, having a high yield of the native and soluble form of the HCV NS3 protease.

These and other objects are achieved using one or more of the embodiments of the present invention described below.

In an embodiment of the invention a procedure is provided for obtaining production of the NS3 serine protease domain in its native form, that is to say containing a bivalent metal ion, which is necessary for the structural integrity of the protein. The innovation in the procedure consists in the addition to the culture medium in which the transformed bacterial cells are grown of compounds containing metals such as Zn, Co, Cd, Mn, Cu, Ni, Ag, Fe, Cr, Hg, Au, Pt, V. These compounds provide the culture medium with the ions required by the protein to take on its native structure. In this way the protein is found in its native, soluble form in the cytoplasm of the bacterial cells, instead of being held in the included bodies, from which it can only be obtained by applying difficult resolubilisation procedures.

In another embodiment of the invention, a procedure is provided that makes it possible to replace the zinc ion in the protease, which is spectroscopically silent, with other ions (for example Co^{2+} or Cd^{2+}), which are spectroscopically active, so as to permit the study of possible inhibitors capable of co-ordinating the metal contained in the protein and therefore of disturbing the bond between the protein and the metal.

In another embodiment of the invention, the addition of bivalent metal ions to a minimum culture medium, containing glucose and ammonium salts enriched with 13C or 15N as the sole sources of carbon and nitrogen, respectively, makes it possible to obtain large amounts

of soluble protein marked with stable heavy isotopes such as 13C or 15N. This type of isotope enrichment is necessary to determine the structure using NMR techniques.

In a further embodiment of the present invention polypeptide sequences are provided that contain the NS3 serine protease domain of hepatitis C virus, suitably modified. These polypeptides are characterised in that they have at their C-terminal end a sequence of extremely hydrophilic amino acids, such as for example a series of lysines, which are not present in the original sequence. By using this other new method there is a substantial improvement in terms of solubility and integrity of the protein produced. These modified protease molecules are also to be considered as a subject of the present invention.

Subjects of the present invention are therefore:

a) Isolated and purified polypeptides containing the HCV NS3 serine protease domain, characterised in that they have at their C-terminal end a tail of at least three lysines.

b) A process for the preparation of polypeptides containing the HCV NS3 serine protease domain in a soluble form, of use for enzymological experiments, determination of the three-dimensional structure of the enzyme both by means of NMR and using X-ray crystallography, comprising the following operations:

- transformation of a prokaryotic host cell with an expression vector containing a DNA sequence coding for a polypeptide with the proteolytic activity of the HCV NS3 protease;

- growth of the prokaryotic host cell on a special culture medium containing Zn^{2+} or alternatively salts of transition metals such as Co, Cd, Mn, Cu, Ni, Ag, Fe, Cr, Hg, Au, Pt, V;

- expression of the DNA sequence required to produce the chosen polypeptide;

- purification of the polypeptide without having to resort to resolubilisation protocols, and without the need for renaturation of the protein from included bodies.

c) A process for the renaturation in vitro of the above polypeptides, characterised in that it comprises the following operations:

- transformation of a prokaryotic host cell with an expression vector containing a DNA sequence coding for a polypeptide with the proteolytic activity of HCV NS3 protease;

- expression of the DNA sequence required to produce the chosen polypeptide;

- purification of the denaturated and renaturated polypeptide of the protein using buffers containing Zn^{2+} or alternatively salts of transition metals such as Co, Cd, Mn, Cu, Ni, Ag, Fe, Cr, Hg, Au, Pt, V.

d) Expression vectors for the production of the polypeptides represented by the sequences SEQ ID NO:1 to SEQ ID NO:4 with the proteolytic activity of HCV NS3, comprising: a polynucleotide coding for one of said polypeptides; regulation, transcription and translation sequences, operating in said host cell, operationally bonded to said polynucleotide; and, optionally, a selectable marker.

e) A prokaryotic cell transformed with an expression vector containing a DNA sequence coding for polypeptides with the proteolytic activity of the HCV NS3 protease, so as to allow said host cell to express the specific polypeptide which is coded in the chosen sequence.

Figure 1 shows the alignment between the HCV NS3 serine protease sequence and the viruses GBV-A, GBV-B and GBV-C/HGV (Hcv, Hga, Hgb, Hgc), with the poliovirus (Pol) 2A cysteine protease. Amino acids conserved in the HCV proteases and in the viruses GBV-A, GBV-B and GBV-C/HGV are shaded. The catalytic residues are underlined and

the residues that bind zinc are indicated using the symbol _.

Figure 2 shows a diagrammatic model of the NS3 serine protease domain. In particular it shows the position within the structure of the amino acids involved in binding zinc (dark grey) and the catalytic triad (light grey).

Figure 3 shows the effect of the zinc ion on HCV NS3 serine protease activity.

Figure 4 shows the effects of the zinc ion on the production of HCV NS3 protease as a soluble protein in E. coli on a minimum culture medium. Column 2 refers to the results of the experiment carried out on the cells without inducing protease production (-IPTG). Columns 3, 4 and 5 indicate that in the absence of $ZnCl_2$ and following the induction of protease production (+IPTG) the protein remains locked in the insoluble portion (indicated by the abbreviation PT). On the contrary, in the presence of $ZnCl_2$ the protease is found entirely in the soluble portion (indicated by the abbreviation SN).

Figure 5 shows the electronic spectrums of the HCV NS3 protease. Figure 5a shows the visible and near-UV spectrum of the Co^{2+}-protease. Figure 5b shows the UV absorption spectrums of the Zn^{2+}-protease and of the Cd^{2+}-protease.

DEPOSITS

Strains of E. coli DH1/p bacteria transformed with the plasmids pT7-7(Pro BK-as K4), pT7-7(Pro)-asK4), pT7-7(Pro H-asK4) and pT7-7 (Pro J8-asK4) and coding for the amino acid sequences SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3 and SEQ ID NO:4, respectively, were deposited on August 8, 1996 with The National Collections of Industrial and Marine Bacteria Ltd (NCIMB), Aberdeen, Scotland, U.K., under access numbers NCIMB 40821, NCIMB 40822, NCIMB 40823 and NCIMB 40824, respectively.

Up to this point a general description has been given of the present invention. With the aid of the

following examples a more detailed description of specific embodiments of the invention will now be given, with the aim of clarifying the objects, characteristics, advantages and methods of application thereof.

EXAMPLE 1

EXPRESSION AND PURIFICATION OF POLYPEPTIDES WITH THE PROTEASE ACTIVITY OF HCV NS3, IN THEIR NATIVE SOLUBLE FORM

The plasmids pT7-7(Pro BK-asK4), pT7-7(Pro H-asK4), pT7-7(Pro J-asK4) and pT7-7(Pro J8-asK4) were constructed to allow expression in E. coli of polypeptides characterised in that they have a sequence chosen from the ones in the group from SEQ ID NO:1 to SEQ ID NO:4. The polypeptides contain the NS3 protease domain of various HCV isolates (BK, H, J and J8, respectively) with the addition of a "tail" of four lysines at the C-terminal end.

pT7-7 (Pro BK-asK4) contains the sequence for HCV-BK (EMBL data bank access number: M58335) between the nucleotides 3411 and 3950, cloned in the vector pT7-7.

pT7-7 (Pro H-asK4) contains the sequence for HCV-H (EMBL data bank access number: M67463) between the nucleotides 3420 and 3959, cloned in the vector pT7-7.

pT7-7 (Pro J-asK4) contains the sequence for HCV-J (EMBL data bank access number: D90208) between the nucleotides 3408 and 3947, cloned in the vector pT7-7.

pT7-7 (Pro J8-asK4) contains the sequence for HCV-J8 (EMBL data bank access number: D10988/D01221) between the nucleotides 3432 and 3971, cloned in the vector pT7-7.

The expression vector pT7-7 is a derivative of pBR322 which contains, in addition to the gene for β-lactamase and the replication origin of ColE1, the promotor and the ribosome binding site of the T7 bacteriophage Ø10 gene (24).

The fragments coding for the HCV NS3 protease were cloned downstream of the T7 bacteriophage Ø10 promoter,

in reading frame with the first ATG condon of the gene 10 protein of phage T7 using methods known to the art.

The cDNA fragment containing the sequence HCV-BK between nucleotides 3411 and 3950 was amplified by Polymerase Chain Reaction (PCR), using the oligonucleotides PROT(BK-K4)S (SEQ ID NO:5) and PROT(BK-K4)AS (SEQ ID NO:6) as primers. The cDNA fragment so obtained was digested with the restriction enzyme *Nde*I, and cloned in pT7-7, which was first linearised with the restriction enzymes *Nde*I and *Sma*I.

The cDNA fragment containing the sequence HCV-H between nucleotides 3420 and 3959 was amplified by PCR, using the oligonucleotides PROT(H-K4)S (SEQ ID NO:7) and PROT(H-K4)AS (SEQ ID NO:8) as primers. The cDNA fragment so obtained was digested with the restriction enzymes *Nde*I and *Eco*RI, and cloned in pT7-7, which was first linearised with the same restriction enzymes.

The cDNA fragment containing the sequence HCV-J between nucleotides 3408 and 3947 was amplified by PCR, using the oligonucleotides PROT(J-K4)S (SEQ ID NO:9) and PROT(J-K4)AS (SEQ ID NO:10) as primers. The cDNA fragment so obtained was digested with the restriction enzymes *Nde*I and *Eco*RI, and cloned in pT7-7, which was first linearised with the same restriction enzymes.

The cDNA fragment containing the sequence HCV-J8 between nucleotides 3432 and 3971 was amplified by PCR, using the oligonucleotides PROT(J8-K4)S (SEQ ID NO:11) and PROT(J8-K4)AS (SEQ ID NO:12) as primers. The cDNA fragment so obtained was digested with the restriction enzymes *Nde*I and *Eco*RI, and cloned in pT7-7, which was first linearised with the same restriction enzymes.

The plasmids pT7-7(Pro BK-asK4), pT7-7(Pro H-asK4), pT7-7(Pro J-asK4) and pT7-7(Pro J8-asK4) containing NS3 sequences also contain the gene for β-lactamase, which can be used as a selection marker for E. coli cells transformed with these plasmids.

The fragments were cloned downstream of the T7 bacteriophage promotor, in reading frame with the first ATG codon of the gene 10 protein of phage T7 using methods known to the art. The plasmids pT7-7(Pro BK-asK4), pT7-7(Pro H-asK4), pT7-7(Pro J-asK4) and pT7-7(Pro J8-asK4) containing NS3 sequences also contain the gene for β-lactamase, which can be used as a selection marker for E. coli cells transformed with these plasmids.

The plasmids are then transformed in the E. coli strain BL21 (DE3), normally used for high levels of expression of genes cloned in expression vectors containing the T7 promotor. In this strain the T7 polymerase gene is carried into the bacteriophage λ DE3, which is integrated into the chromosome of BL21 cells (25). Expression of the gene is induced by incubating the cultures at an A600 nm of 0.7-0.9 with 0.4 mM of isopropyl-1-thio-β-D-galactopyranoside (IPTG) for 3 hours at 20°C in LB culture medium additioned with ZnCl2 at a concentration that can vary from 50 μM to 1 mM. After the three hours have passed the cells are harvested and washed in a saline phosphate buffer solution (20 mM sodium phosphate pH 7.5, 140 mM NaCl), after which they are re-suspended in 25 mM sodium phosphate at pH 7.5, 10% glycerol, 500 mM NaCl, 10 mM DTT, 0.5% CHAPS (10 ml per 1 litre of culture medium). The cells are then lysated by passing twice through a "French pressure cell" and the homogenate obtained in this way is centrifugated at 100,000xg for 1 hour, while the nucleic acids are removed by precipitation with 0.5% polyethylenimine. The supernatants are loaded onto a HiLoad 16/10 SP Sepharose High Performance column (Pharmacia), and balanced with 50 mM of sodium phosphate at pH 7.5, 5% glycerol, 3 mM DTT, 0.1% CHAPS (buffer A). The column had been washed repeatedly with buffer A and the protease was eluted by applying a gradient of from 0 to 0.6 M NaCl. The fractions containing the protease were then collected and concentrated using a chamber for ultrafiltration under

magnetic stirring, equipped with a YM-10 membrane (Amicon). The sample was then loaded onto an HR 26/60 HiLoad Superdex 75 column (Pharmacia), balanced with buffer A, operating at a flow rate of 1 ml/min.

The fractions containing NS3 were collected and further purified on an HR 5/5 Mono S column (Pharmacia), balanced with buffer B and operating at a flow rate of 1 ml/min. The protease was eluted from the column in pure form applying a linear gradient of 0-0.6 NaCl in buffer A.

After this passage the protein was preserved in stocks at concentrations of 50-150 μM at a temperature of -80 °C after freezing in liquid nitrogen. The concentration of the protein was estimated by determination of absorbancy at 280 nm using a coefficient of extinction deriving from the sequence data or from quantitative amino acid analysis. Both methods come to the same results, with an error factor of 10%. The purity of the enzyme was ascertained on SDS polyacrylamide gel and by HPLC using an inverse phase Vydac C4 column (4.6x250 mm, 5 mm, 300 A). The eluents used were H2O/0.1% TFA (A) and acetonitryl/0.1% TFA (B). A linear gradient of from 3% to 95% B over 60 minutes was used. Analysis of the N-terminal end was carried out using Edman degradation on a gaseous phase sequencer (Applied Biosystem model 470A) and the analysis by mass spectroscopy revealed that more than 96% of the purified protein has the N-terminal sequence PITAYSSQ. The remaining 3% has the sequence MAPITAYSSQ as foreseen from the data on the nucleotide sequence.

In order to measure the enzymatic activity of the purified protein, a synthetic peptide of 13 amino acids was used as a substrate. This peptide was derived from the cleavage sequence of the NS4A-NS4B junction (DEEMECSSHLPYK). A peptide with 14 amino acids corresponding to the central hydrophobic region of the

protein NS4A (from position 21 to position 34) (Pep4A21-34: GSVVIVGRIILSGR) was used as a protease cofactor.

The peptides were synthesised by solid phase synthesis based on Fmoc chemistry. After washing and deprotection, the "raw" peptides were purified by HPLC to 98% purity. The identity of the peptides was determined by mass spectrometry. The peptide solutions stored were prepared in DMSO and preserved at -80°C, furthermore the concentrations were determined by quantitative amino acid analysis carried out on samples hydrolysed with HCl.

The cleavage tests were carried out using 300 nM - 1.6 µM of enzyme in 30 l of 50 mM Tris pH 7.5, 50% glycerol, 2% CHAPS, 30 mM DTT and appropriate amounts of substrate and/or peptide-NS4A at 22°C. The reaction was stopped by addition of 70µl of H2O containing 0.1% TFA. Cleavage of the peptide substrate was determined by HPLC using a Merck-Hitachi chromatograph. After this, 90µl of each sample were injected into an inverse phase Lichrospher C18 cartridge column (4x125 mm, 5µm, Merck) and the fragments were separated using an acetonitryl gradient of 3-100% at 2%/min. Identification of the peak was achieved following both the absorbancy at 220 nm and the fluorescence of the tyrosine (λex= 260 nm, λem= 305 nm).

Tables 1 and 2 give the data for solubility and yield relating to the NS3 protease corresponding to various HCV virus isolated. Table 1 gives the data for production of the various forms of protease both with and without the addition of four lysines at the C-terminal end, and both with and without the addition of ZnCl2 in the culture medium. The data are expressed as the percentage of protein recovered in the soluble fraction of the cell extracts and the protein found in the included bodies. Table 2 gives the yields and solubility of the various forms of protease, purified from E. coli cells grown in the presence of ZnCl2. As can be seen from the results given, the modified proteases (BK-ASK4,

J-ASK4, H-ASK4) are between 10 and 20 times more soluble and, when expressed in a culture medium containing an excess of ZnCl2, they give a yield up to 10 times greater than the respective proteases without the lysine tail.

TABLE 1

Construct	Culture medium	Soluble portion	Included bodies
Pro J	LB	5%	95%
Pro J-asK4	LB	20%	80%
Pro J	LB + $ZnCl_2$	99%	<1%
Pro J-asK4	LB + $ZnCl_2$	99%	<1%
Pro H	LB	<2%	>98%
Pro H-asK4	LB	3-4%	>95%
Pro H	LB + $ZnCl_2$	5%	95%
Pro H-asK4	LB + $ZnCl_2$	50%	50%

TABLE 2

Construct	Yield (mg/lt medium)	Solubility (mg/ml)
Pro BK	1-2	1-2
Pro BK-asK4	10-15	>40
Pro H	0.1-0.2	1-2
Pro H-asK4	1-2	>40
Pro J	1-2	0.5-1
Pro J-asK4	15-20	>10

EXAMPLE 2

DETERMINATION OF THE METAL CONTENT OF POLYPEPTIDES WITH THE PROTELYTIC ACTIVITY OF HCV NS3 PROTEASE

The polypeptides purified according to the procedure described in examples 1, 3 and 5 were further dialysed against buffers containing a chelating agent, in order to remove any metal ions bound to the protein, and their metal content was determined by atomic absorption spectrometry using a Perkin-Elmer Instrument

spectrometer. The glass equipment used for analysis of the metal content was washed using 30% nitric acid and rinsed completely with deionised water. The protease (at a concentration of 4 mg/ml) was dialysed for a period of at least 16 hours against a buffer containing 50 mM Tris/Hcl pH 7.5, 3 mM DTT, 10% glycerol, 0.1% CHAPS. A Chelex-100 resin (2.5 g/l) was held in suspension in the dialysis buffer to prevent contamination by casual metal ions. The protein was then hydrolysed with nitric acid and then used to determine the metal content. The standardised Zn^{2+}, Co^{2+} and Cd^{2+} solutions were purchased from Merck.

The metal content was found to be 1 g-atom per 1 mole enzyme (see table 3 - n.d.= not determined), with the exception of of the apoprotein, which has a negligible metal content.

TABLE 3

Protein	Zn (g-atoms/mole)	Co (g-atoms/mole)	Cd (g-atoms/mole)
Zn^{2+}-NS3	1.09	n.d.	n.d.
Apo-NS3	0.02	n.d.	n.d.
Co^{2+}-NS3	0.19	0.90	n.d.
Cd^{2+}-NS3	0.09	n.d.	1.15

n.d.: not determined

EXAMPLE 3

PROCEDURE FOR THE RENATURATION OF THE NS3 PROTEASE IN THE PRESENCE OF ZINC

To ascertain whether or not zinc is required for HCV NS3 serine protease activity, its proteolytic activity was first measured on a synthetic substrate peptide. This measurement was carried out in the presence of increasing concentrations of EDTA or of 1,10-

phenanthroline. It was found that these two compounds do not inhibit proteolysis by NS3 at concentrations lower than 1 mM. Above these concentrations both EDTA and 1,10-phenanthroline only show a modest level of inhibition of NS3 activity. However a similar inhibition behaviour has been obtained in control experiments using structurally similar elements to 1,10-phenanthroline, which is not capable of chelating zinc ions, and the activity was not re-obtained in the presence of an excess of Zn^{2+} ions. These results suggest that either zinc is not required for enzymatic activity, or that it is so strongly bonded to the protein that it cannot be removed by treatment with chelating agents. It was therefore decided to proceed with preparation of a protein containing no zinc (apoprotein) and to measure its biochemical activity in the absence and in the presence of this metal. Bonded zinc cannot be removed by dialysis against chelators with a pH exceeding 7, whereas on the other hand prolonged dialysis of the enzyme at a pH of less than 5 and in the presence of 10 mM EDTA causes a loss of zinc accompanied by irreversible precipitation of the sample. The above observations suggest that the zinc is strongly bound and that it is essential for the structural integrity of the protein. In order to facilitate the release of zinc the apoprotein was obtained by applying the following procedure: 1.7 mg of NS3 protease were denaturated by addition of TFA to a final concentration of 1%. The denaturated protein was then purified on a Resouce RPC 3 ml column using an acetonitryl gradient of from 0% to 85% in the presence of 0.1% TFA. The flow rate of the column was equivalent to 2 ml/min and the volume of the gradient was 45 ml. The zinc content of the apoprotein was found to be negligible. The enzymatic activity of the apoprotein was then tested in the presence and in the absence of zinc. The apoprotein was diluted to a final concentration of 60 nM in the activity buffer containing the concentrations

of $ZnCl_2$ shown in the graph and 10 mM DTT to prevent oxidation of the thiole groups. After an incubation period of 1 hour at 22°C the reaction was started by adding the substrate eptide at a concentration of 40 mM. The reaction was then made to proceed for another hour before taking the measurements. As shown in figure 3, reconstitution of the enzymatic activity depends on the concentration of zinc ions in the buffer. Maximum reactivation was observed at a ZnCl2 concentration of 25 µM. At this concentration the enzymatic activity is found to be approximately 50% when compared to the protease containing zinc (diluted in the same buffer at the same final concentration). This experiment gives unequivocal proof that zinc is necessary in order for the enzyme to be structurally complete and active, and it also provides a method for reconstitution of NS3 serine protease activity starting from the apoprotein.

EXAMPLE 4

PROCESS FOR THE PRODUCTION OF HCV NS3 PROTEASE IN A FORM THAT CAN BE USED FOR DETERMINATION OF THE THREE-DIMENSIONAL STRUCTURE THEREOF USING NMR TECHNIQUES

The discovery that HCV NS3 protease contains a structural zinc atom has been used to increase the production of soluble protein in bacterial cells (E. coli) and therefore to produce a protein in a form that can be used for experiments aimed at determining the structure by means of NMR.

In effect, determination of structure by means of NMR involves metabolic marking with 15_N and 13_C, to be carried out on a minimum culture medium, for example modified M9 culture medium $(NH_4)_2SO_4$ 1g/l, K-phosphate 100 mM, $MgSO_4$ 0.5 mM, $CaCl_2$ 0.5 mM, biotin 5µM, thiamine 7µM, ampicillin 5µg/ml, glucose 4 g/l, $FeSO_4.7H_2O$ 13 µM). Induction in this culture medium, which does not include zinc salts in its composition, inevitably results in the production of insoluble protein, whereas the addition of 50µM of $ZnCl_2$ results in the production of a completely

soluble protease. In this way it is possible to produce a marked protein using $(^{15}NH_4)_2SO_4$ as a source of nitrogen and ^{13}C-glucose as a source of carbon.

Following this new procedure, a protein has been obtained that remains in a soluble form in the cytoplasm and is not captured by the inclusion bodies, as was the case using the old procedures. In this way the resolubilisation procedures become unnecessary, which results in considerable advantages, as these procedures have an extremely variable yield, require extremely controlled conditions and also frequently cause irreversible alterations in the protein. Figure 4 shows how the protease (at approximately 21 kDa - indicated in the figure by an arrow) is produced as an insoluble aggregate (PT) when the bacterial cells are grown in minimum culture medium without zinc (columns 3, 4 and 5). On the contrary, if $ZnCl_2$ is added to the culture medium at a concentration of 50 mM the protein is found in the soluble fraction (SN) (columns 6, 7 and 8) and disappears from the insoluble fraction (PT).

EXAMPLE 5

REPLACEMENT OF THE Zn^{2+} BOUND TO NS3 WITH SPECTROSCOPIC PROBES SUCH AS Co^{2+} OR Cd^{2+}

The Zn^{2+} binding site of the HCV NS3 protease and zinc can be studied by replacing the zinc with metals that make spectroscopic studies possible. The close binding of the structural zinc to the enzyme makes it difficult to remove the metal and replace it *in vitro*. As a result, the Zn^{2+} was replaced by Co^{2+} and Cd^{2+} by incorporation in vivo. The bacterial cells (E. coli) were transformed with an appropriate expression vector and grown in minimum culture medium containing 100 mM potassium phosphate at pH 7.0, 0.5 mM $MgSO_4$, 0.5 mM $CaCl_2$, 13 µM $FeSO_4$, 7 µM thiamine, 6 µM biotin. Glucose (4 g/l) and (NH4)2SO4 (1 g/l) were used as sources of carbon and nitrogen, respectively. To reduce the amount of Zn^{2+} in the culture medium, the phosphate buffer was

made to pass through a Chelex-100 column. To obtain production of Co^{2+} or Cd^{2+}-NS3, 50 mM of CoCl2 and CdCl2 were added, respectively, 20 minutes before addition of IPTG. Purification of the Co^{2+} and Cd^{2+}-proteases was obtained using the procedure described in example 1, except for the fact that all the buffers used were treated with Chelex-100 resin (2.5 g/l) and the DTT was eliminated.

The addition of CoCl2 or CdCl2 to the culture medium still results in production of the soluble enzyme, which indicates that the Co^{2+} and Cd^{2+} ions can replace zinc in the binding site for metal and protease.

The protease containing Co^{2+} and Cd^{2+} was subjected to electronic absorption spectroscopic analysis. The protease containing Co^{2+} shows a typical absorption spectrum in the visible region (figure 6a), which indicates a binding site with a tetrahedral geometry (26). The two main bands at 640 nm and at 685 nm and the minimums at 585 nm and at 740 nm indicate *d-d* transitions. The energy in these transitions and the molar extinction coefficients are characteristic of complexes with a distorted tetrahedral co-ordination geometry (27). The *d-d* transition energy is consistent, with a mixed sulphur-nitrogen co-ordination bond. Furthermore, the centroide in the band corresponding to the *d-d* transition indicates a Co^{2+} complex with a S3N bond (26). A typical charge transfer band S -> Co^{2+} was observed at around 365 nm (figure 6a), implying that the metal ion is co-ordinated by thiolates.

In accordance with these data, the UV absorbancy spectrum of the Cd^{2+}-protease (figure 6b) shows an increase in absorbancy at around 250 nm, which in all probability is due to a charge transfer band S -> Cd^{2+} (28). In conclusion, spectroscopic analysis of the Co^{2+} and Cd^{2+}- proteases is completely consistent with the three-dimensional model proposed by us. In face, in the model the binding site for the metal is made up of three

thiole groups of three cysteines and of a nitrogen atom from the side chain of a hystidine. Each of the residues that according to the model form the binding site for the metal has been changed to alanine and, as expected, none of the mutants obtained is capable of being expressed in a soluble form in E. coli.

BIBLIOGRAPHY

1. Behrens S.E., Tomei L., De Francesco R.(1996) EMBO J.15:12-22.

2. Hijikata M., Kato N., Ootuyama Y., Nakagawa M. & Shimotohno K. (1991) Proc. Natl. Acad. Sci. USA 88:5547-5551.

3. Grakoui A., McCourt D.W., Wychowski C., Feinstone S.M., Rice C.M. (1993) Proc. Natl. Acad. Sci. USA 90: 10583-10587.

4. Bartenschlager R., Ahlborn-Laake L., Mous J. & Jacobsen H. (1993) J. Virol. 68:1045-1055.

5. Eckart M.R., Selby M., Masiarz F., Lee C., Berger K., Crawford K., Kuo C., Kuo G., Houghton M. & Choo Q.-L. (1993) Biochem. Biophys. Res. Comm. 192:399-406.

6. Grakoui A., McCourt D.W., Wychowski C., Feinstone S.M. & Rice C.M. (1993) J. Virol. 67:2832-2843.

7. Tomei L., Failla C., Santolini E., De Francesco R. & La Monica N. (1993) J. Virol. 67:4017-4026.

8. Manabe S., Fuke I., Tanishita O., Kaji C., Gomi Y., Yoshida S., Mori C., Takamizawa A., Yohida I. & Okayama H. (1994) virology 198:636-644.

9. Pizzi E., Tramontano A., Tomei L., La Monica N., Failla C., Sardana M., Wood T., De Francesco R. (1994) Proc. Natl. Acad. Sci. USA 91:888-892.

10. Shimuzu Y., Yamanii K., Masuho Y., Yokota T., Inouhe H., Sudo K., Satoh S. & Shimothono K. (1996) J. Virol. 70:127-132.

11. Lin C., Thomson J.A. & Rice C.M (1995) J. Virol. 69:4373-4380.

12. Tomei L., Failla C., Vitale R.L., Bianchi E. & De Francesco R. (1996) J. Gen. Virol. 77:1065-1070.

13. Chambers T.J., Weir R.C., Grakoui A., McCourt D.W., Bazan J.F., Fletterick R.J., Rice C.M. (1990) Proc. Natl. Acad. Sci. USA 87:8898-8902.

14. Lam P.Y., Jadhav P.K., Eyermann C.J., Hodge C.N., Ru Y., Bacheler L.T., Meek J.L., Otto M.J., Rayner M.M., Wong Y.N., Chang C.-H., Weber P.C., Jackson D.A., Sharpe T.R., Erickson-Viitanen S. (1994) Science 263:380-384.

15. Georgiou G. and Valax P. (1996) Current Opinion in Biotechnology 7:190-197.

16. Failla C., Pizzi E., De Francesco R., Tramontano A. (1996) Folding & Design 1:35-42.

17. Zuckermann A.J.(1996) The Lancet 347:58-559.

18. Muerhoff A.S., Leary T.P., Simons J.N., Pilot-Matias T.J., Dawson G.J., Erker J.C., Chalmers M.L., Schlauder G.C., Desai S.M. & Mushahwar I.K. (1995) J. Virol. 69:5621-5630.

19. Leary T.P.. Muerhoff A.S., Simons J.N., Pilot-Matias T.J., Erker J.C., Chalmers M.L., Schlauder G.C., Dawson G.J., Desai S.M. & Mushahwar I.K. (1996) J.Medical Virol. 48:60-67.

20. Yu S.F. and Lloyd R.E. (1992) Virology 186:725-735

21. Voss T., Meyer R. and Sommergruber W. (1995) Protein Science 4:2526-2531.

22. Rice W.G., Schaeffer, C.A., Harten B., Villinger F., South T.L., Summers M.F., Henderson L.E., Bess J.W.J., Arthur L.O., McDougall J.S., Orloff S.L., Mendeleyev J. & Kun E. (1993) 361:473-475.

23. Rice W.G., Supko J.G., Malspeis L., Buckheit R.W.J., Clanton D., Bu M., Graham L., Schaeffer C.A., Turpin J.A., Domogala J., Gogliotti R., Bader J.P., Halliday S.M., Coren L., Sowder R.C.I., Arthur L.O. & Henderson L.E. (1995) Science 270:1194-1197.

24. Tabor S. & Richardson C. C. (1985) Proc. Natl. Acad. Sci. USA 82:1074-1078.

25. Studier F.W. and Moffatt (1986) J.Mol.Biol. 189:113-130

26. Maret W.& Vallee B.L. (1993) Meth. Enzym. 226:52-71.

27. Bertini I.& Luchinat C. (1984) Adv. Inorg. Biochem. 6:72-111.

28. Fitzgerald D.W. and Coleman J.E. (1991) Biochemistry 30:5195-5201.

SEQUENCE LISTING

GENERAL INFORMATION

(i) APPLICANT:

ISTITUTO DI RICERCHE DI BIOLOGIA MOLECOLARE P.ANGELETTI S.p.A.

(ii) TITLE OF INVENTION: "SOLUBLE POLYPEPTIDES WITH ACTIVITY OF THE NS3 SERINE PROTEASE OF HEPATITIS C VIRUS, AND PROCESS FOR THEIR PREPARATION AND ISOLATION"

(iii) NUMBER OF SEQUENCES: 12

(iv) MAILING ADDRESS:

(A) ADDRESSEE: Societa' Italiana Brevetti

(B) STREET: Piazza di Pietra, 39

(C) CITY: Rome

(D) COUNTRY: Italy

(E) POST CODE: I-00186

(v) COMPUTER-READABLE FORM:

(A) TYPE OF SUPPORT: Floppy disk 3.5'' 1.44 MBYTES

(B) COMPUTER: IBM PC compatible

(C) OPERATING SYSTEM: PC-DOS/MS-DOS Rev. 5.0

(D) SOFTWARE: Microsoft Word 6.0

(viii) AGENT INFORMATION

(A) NAME: DI CERBO Mario (Dr.)

(B) REFERENCE: RM/X88878/PC-DC

(ix) TELECOMMUNICATIONS INFORMATION

(A) TELEPHONE: 06/6785941

(B) TELEFAX: 06/6794692

(C) TELEX: 612287 ROPAT

(1) INFORMATION ON SEQUENCE SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS

(A) LENGTH: 187 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ix) FEATURE

(A) NAME: Pro BK-asK4

(D) OTHER INFORMATION: sequence for the NS3 protease of HCV-isolated BK.

(xi) SEQUENCE DESCRIPTION SEQ ID NO: 1:

Met Ala Pro Ile Thr Ala Tyr Ser Gln Gln Thr Arg Gly Leu Leu Gly
1 5 10 15
Cys Ile Ile Thr Ser Leu Thr Gly Arg Asp Lys Asn Gln Val Glu Gly
20 25 30
Glu Val Gln Val Val Ser Thr Ala Thr Gln Ser Phe Leu Ala Thr Cys
35 40 45
Val Asn Gly Val Cys Trp Thr Val Tyr His Gly Ala Gly Ser Lys Thr
50 55 60
Leu Ala Gly Pro Lys Gly Pro Ile Thr Gln Met Tyr Thr Asn Val Asp
65 70 75 80
Gln Asp Leu Val Gly Trp Gln Ala Pro Pro Gly Ala Arg Ser Leu Thr
85 90 95
Pro Cys Thr Cys Gly Ser Ser Asp Leu Tyr Leu Val Thr Arg His Ala
100 105 110
Asp Val Ile Pro Val Arg Arg Arg Gly Asp Ser Arg Gly Ser Leu Leu
115 120 125
Ser Pro Arg Pro Val Ser Tyr Leu Lys Gly Ser Ser Gly Gly Pro Leu
130 135 140
Leu Cys Pro Ser Gly His Ala Val Gly Ile Phe Arg Ala Ala Val Cys
145 150 155 160
Thr Arg Gly Val Ala Lys Ala Val Asp Phe Val Pro Val Glu Ser Met
165 170 175
Glu Thr Thr Met Arg Ala Ser Lys Lys Lys Lys
180 185

(2) INFORMATION ON SEQUENCE SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS

(A) LENGTH: 187 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ix) FEATURE

(A) NAME: Pro H-asK4

(D) OTHER INFORMATION: sequence for the NS3 protease of HCV-isolated H.

(xi) SEQUENCE DESCRIPTION SEQ ID NO: 2:

Met Ala Pro Ile Thr Ala Tyr Ala Gln Gln Thr Arg Gly Leu Leu Gly
1 5 10 15

Cys Ile Ile Thr Ser Leu Thr Gly Arg Asp Lys Asn Gln Val Glu Gly
20 25 30

Glu Val Gln Ile Val Ser Thr Ala Thr Gln Thr Phe Leu Ala Thr Cys
35 40 45

Ile Asn Gly Val Cys Trp Thr Val Tyr His Gly Ala Gly Thr Arg Thr
50 55 60

Ile Ala Ser Pro Lys Gly Pro Val Ile Gln Met Tyr Thr Asn Val Asp
65 70 75 80

Gln Asp Leu Val Gly Trp Pro Ala Pro Gln Gly Ser Arg Ser Leu Thr
85 90 95

Pro Cys Thr Cys Gly Ser Ser Asp Leu Tyr Leu Val Thr Arg His Ala
100 105 110

Asp Val Ile Pro Val Arg Arg Arg Gly Asp Ser Arg Gly Ser Leu Leu
115 120 125

Ser Pro Arg Pro Ile Ser Tyr Leu Lys Gly Ser Ser Gly Gly Pro Leu
130 135 140

Leu Cys Pro Ala Gly His Ala Val Gly Leu Phe Arg Ala Ala Val Cys
145 150 155 160

Thr Arg Gly Val Ala Lys Ala Val Asp Phe Ile Pro Val Glu Asn Leu
165 170 175

Glu Thr Thr Met Arg Ala Ser Lys Lys Lys Lys
180 185

(3) INFORMATION ON SEQUENCE SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS

(A) LENGTH: 187 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ix) FEATURE

(A) NAME: ProJ-asK4

(D) OTHER INFORMATION: sequence for the NS3 protease of HCV-isolated J.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

Met Ala Pro Ile Thr Ala Tyr S r Gln Gln Thr Arg Gly Leu Leu Gly
1 5 10 15
Cys Ile Ile Thr Ser Leu Thr Gly Arg Asp Lys Asn Gln Val Asp Gly
20 25 30
Glu Val Gln Val Leu Ser Thr Ala Thr Gln Ser Phe Leu Ala Thr Cys
35 40 45
Val Asn Gly Val Cys Trp Thr Val Tyr His Gly Ala Gly Ser Lys Thr
50 55 60
Leu Ala Gly Pro Lys Gly Pro Ile Thr Gln Met Tyr Thr Asn Val Asp
65 70 75 80
Gln Asp Leu Val Gly Trp Pro Ala Pro Pro Gly Ala Arg Ser Met Thr
85 90 95
Pro Cys Thr Cys Gly Ser Ser Asp Leu Tyr Leu Val Thr Arg His Ala
100 105 110
Asp Val Val Pro Val Arg Arg Arg Gly Asp Ser Arg Gly Ser Leu Leu
115 120 125
Ser Pro Arg Pro Ile Ser Tyr Leu Lys Gly Ser Ser Gly Gly Pro Leu
130 135 140
Leu Cys Pro Ser Gly His Val Val Gly Ile Phe Arg Ala Ala Val Cys
145 150 155 160
Thr Arg Gly Val Ala Lys Ala Val Asp Phe Ile Pro Val Glu Ser Met
165 170 175
Glu Thr Thr Met Arg Ala Ser Lys Lys Lys Lys
180 185

(4) INFORMATION ON SEQUENCE SEQ ID NO 4:

(i) SEQUENCE CHARACTERISTICS

(A) LENGTH: 186 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ix) FEATURE

(A) NAME: Pro J8-asK4

(D) OTHER INFORMATION: sequence for the NS3 protease of HCV-isolated J8.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

Ala Pro Ile Thr Ala Tyr Thr Gln Gln Thr Arg Gly Leu Leu Gly Ala
1 5 10 15

Ile Val Val Ser Leu Thr Gly Arg Asp Lys Asn Glu Gln Ala Gly Gln
20 25 30
Val Gln Val Leu Ser Ser Val Thr Gln Thr Phe Leu Gly Thr Ser Ile
35 40 45
Ser Gly Val Leu Trp Thr Val Tyr His Gly Ala Gly Asn Lys Thr Leu
50 55 60
Ala Gly Pro Lys Gly Pro Val Thr Gln Met Tyr Thr Ser Ala Glu Gly
65 70 75 80
Asp Leu Val Gly Trp Pro Ser Pro Pro Gly Thr Lys Ser Leu Asp Pro
85 90 95
Cys Thr Cys Gly Ala Val Asp Leu Tyr Leu Val Thr Arg Asn Ala Asp
100 105 110
Val Ile Pro Val Arg Arg Lys Asp Asp Arg Arg Gly Ala Leu Leu Ser
115 120 125
Pro Arg Pro Leu Ser Thr Leu Lys Gly Ser Ser Gly Gly Pro Val Leu
130 135 140
Cys Ser Arg Gly His Ala Val Gly Leu Phe Arg Ala Ala Val Cys Ala
145 150 155 160
Arg Gly Val Ala Lys Ser Ile Asp Phe Ile Pro Val Glu Ser Leu Asp
165 170 175
Val Ala Thr Arg Ala Ser Lys Lys Lys Lys
180 185

(5) INFORMATION ON SEQUENCE SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS

(A) LENGTH: 26 nucleotides

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Synthetic DNA

(iv) ANTISENSE: No

(vii) IMMEDIATE SOURCE: oligonucleotide synthesiser

(ix) FEATURE

(A) NAME: PROT(BK-K4)S

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

GCATACATAT GGCGCCCATC ACGGCC 26

(6) INFORMATION ON SEQUENCE SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS

(A) LENGTH: 33 nucleotides

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) ASPECT: linear

(ii) MOLECULE TYPE: Synthetic DNA

(iv) ANTISENSE: Yes

(vii) IMMEDIATE SOURCE: oligonucleotide synthesiser

(ix) FURTHER CHARACTERISTICS

(A) NAME: PROT(BK-K4)AS

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

CTACTTCTTC TTCTTGCTAG CCCGCATAGT AGT 33

(7) INFORMATION ON SEQUENCE SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS

(A) LENGTH: 26 nucleotides

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Synthetic DNA

(iv) ANTISENSE: No

(vii) IMMEDIATE SOURCE: oligonucleotide synthesiser

(ix) FEATURE

(A) NAME: PROT(H-K4)S

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

GAGATACATA TGGCGCCTAT CACGGC 26

(8) INFORMATION ON SEQUENCE SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS

(A) LENGTH: 42 nucleotides

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Synthetic DNA

(iv) ANTISENSE: Yes

(vii) IMMEDIATE SOURCE: oligonucleotide synthesiser

(ix) FEATURE

(A) NAME: PROT(H-K4)AS

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

TTTGAATTCC TACTTCTTCT TCTTGCTAGC TCTCATGGTT GT 42

(9) INFORMATION ON SEQUENCE SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS

(A) LENGTH: 27 nucleotides

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Synthetic DNA

(iv) ANTISENSE: No

(vii) IMMEDIATE SOURCE: oligonucleotide synthesiser

(ix) FEATURE

(A) NAME: PROT(J-K4)S

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

TTTCATATGG CGCCTATCAC GGCCTAT 27

(10) INFORMATION ON SEQUENCE SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS

(A) LENGTH: 26 nucleotides

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Synthetic DNA

(iv) ANTISENSE: Yes

(vii) IMMEDIATE SOURCE: oligonucleotide synthesiser

(ix) FEATURE

(A) NAME: PROT(J-K4)AS

(xi) SEQUENCE DESCRIPTION SEQ ID NO: 10:

TTTGAATTCC TACTTCTTCT TCTTGCTAGC CCGCATGGTA GT 42

(11) INFORMATION ON SEQUENCE SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS

(A) LENGTH: 24 nucleotides

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: Synthetic DNA
(iv) ANTISENSE: No
(vii) IMMEDIATE SOURCE: oligonucleotide synthesiser
(ix) FEATURE
(A) NAME: PROT(J8-K4)S
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

GGGAATTCCA TATGGCTCCC ATTACTGCT ACAC 24

(12) INFORMATION ON SEQUENCE SEQ ID NO: 12:
(i) SEQUENCE CHARACTERISTICS
(A) LENGTH: 42 nucleotides
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: Synthetic DNA
(iv) ANTISENSE: Yes
(vii) IMMEDIATE SOURCE: oligonucleotide synthesiser
(ix) FEATURE
(A) NAME: PROT(J8-K4)S
(xi) SEQUENCE DESCRIPTION SEQ ID NO: 12:

TTTGAATTCC TACTTCTTCT TCTTGCTAGC CCGTGTGGCG AC 42

CLAIMS

1. Isolated and purified polypeptides containing the HCV NS3 serine protease domain sequence, characterised in that they have at their C-terminal end a tail of at least three lysines.

2. Expression vectors for the production of the polypeptides according to claim 1, characterised in that they comprise: a polynucleotide coding for one of said polypeptides; regulation and translation sequences functional in said host cell, operationally bonded to said polynucleotide; and, optionally, a selectable marker.

3. A prokaryotic cell, characterised in that it is transformed with an expression vector containing a DNA sequence coding for the polypeptides according to claim 1, so as to allow said host cell to express the specific polypeptide which is coded in the chosen sequence.

4. A process for the preparation of polypeptides containing the HCV NS3 serine protease domain sequence, characterised in that it comprises the following operations:

- transformation of a prokaryotic host cell with an expression vector containing a DNA sequence coding for a polypeptide containing the HCV NS3 serine protease domain sequence;

- growth of the prokaryotic host cell on a special culture medium containing Zn^{2+} or alternatively salts of transition metals such as Co, Cd, Mn, Cu, Ni, Ag, Fe, Cr, Hg, Au, Pt, V;

- expression of the DNA sequence required to produce the chosen polypeptide;

- purification of the polypeptide without having to resort to resolubilisation protocols, and without the need for renaturation of the protein from included bodies,

said procedure making it possible to obtain said polypeptides in their native, soluble form suitable to

enable determination of the three-dimensional structure of the enzyme by means of NMR or X-ray crystallography techniques.

5. A process for the renaturation *in vitro* of polypeptides containing the HCV NS3 serine protease domain sequence, characterised in that it comprises the following operations:

- transformation of a prokaryotic host cell using an expression vector containing a DNA sequence coding for a polypeptide that contains the HCV NS3 serine protease domain sequence;

- expression of the DNA sequence required to produce the chosen polypeptide;

- purification of the denaturated polypeptide and renaturation of the protein using buffers containing Zn^{2+} or alternatively salts of transition metals such as Co, Cd, Mn, Cu, Ni, Ag, Fe, Cr, Hg, Au, Pt, V,

said procedure making it possible to obtain said polypeptides in their native, soluble form suitable to enable determination of the three-dimensional structure of the enzyme by means of NMR or X-ray crystallography techniques.

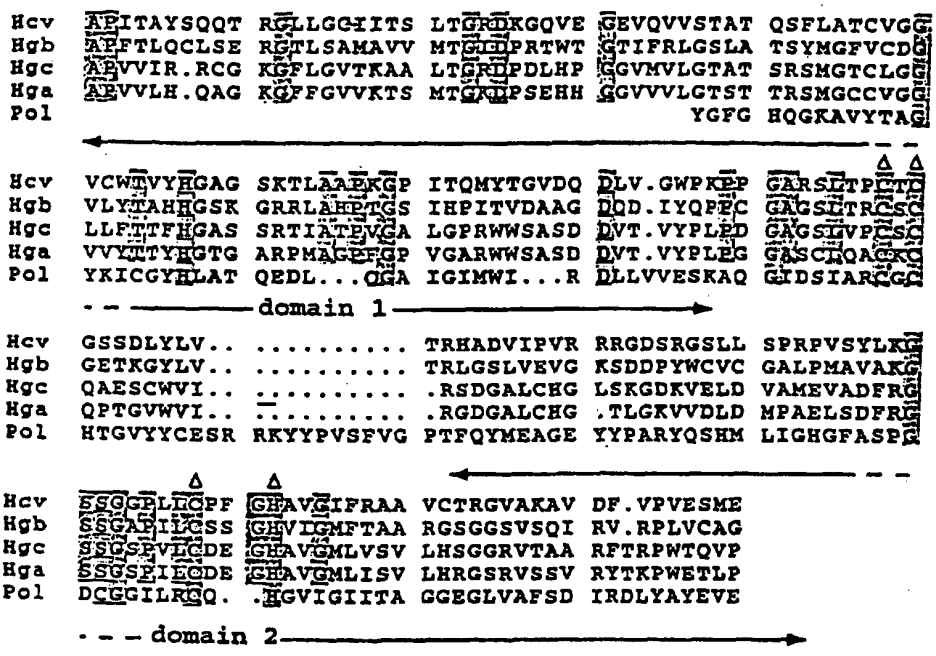
Hcv APITAYSQQT RGLLGCIITS LTGRDKGQVE GEVQVVSTAT QSFLATCVGG
Hgb APFTLQCLSE RGTLSAMAVV MTGCDPRTWT GTIFRLGSLA TSYMGFVCDG
Hgc APVVIR.RCG KGFLGVTKAA LTGRDPDLHP GGVMVLGTAT SRSMGTCLGG
Hga APVVLH.QAG KGFFGVVKTS MTGKDPSEHH GGVVVLGTST TRSMGCCVGG
Pol YGFG HQGKAVYTAG
Hcv VCWTVYHGAG SKTLAAPKGP ITQMYTGVDQ DLV.GWPKPP GARSLTPCTC
Hgb VLYTAHHGSK GRRLAHPTGS IHPITVDAAG DQD.IYQPFC GAGSLTRCSC
Hgc LLFTTFHGAS SRTIATPVGA LGPRWWSASD DVT.VYPLPD GAGSLVPCSC
Hga VVYTTYHGTG ARPMAGPFGP VGARWWSASD DVT.VYPLEG GASCLQACKC
Pol YKICGYHLAT QEDL...CGA IGIMWI...R DLLVVESKAQ GIDSIARCGC
domain 1
Hcv GSSDLYLV.. ..... TRHADVIPVR RRGDSRGSLL SPRPVSYLKG
Hgb GETKGYLV.. ..... TRLGSLVEVG KSDDPYWCVC GALPMAVAKG
Hgc QAESCWVI.. ..... .RSDGALCHG LSKGDKVELD VAMEVADFRG
Hga QPTGVWVI.. ..... .RGDGALCHG .TLGKVVDLD MPAELSDFRG
Pol HTGVYYCESR RKYYPVSFVG PTFQYMEAGE YYPARYQSHM LIGHGFASPG
Hcv SSGGPLLCPF GHAVGIFRAA VCTRGVAKAV DF.VPVESME
Hgb SSGAPILCSS GHVIGMFTAA RGSGGSVSQI RV.RPLVCAG
Hgc SSGSPVLCDE GHAVGMLVSV LHSGGRVTAA RFTRPWTQVP
Hga SSGSPILCDE GHAVGMLISV LHRGSRVSSV RYTKPWETLP
Pol DCGGILRCQ. .HGVIGIITA GGEGLVAFSD IRDLYAYEVE
domain 2

FIG. 1

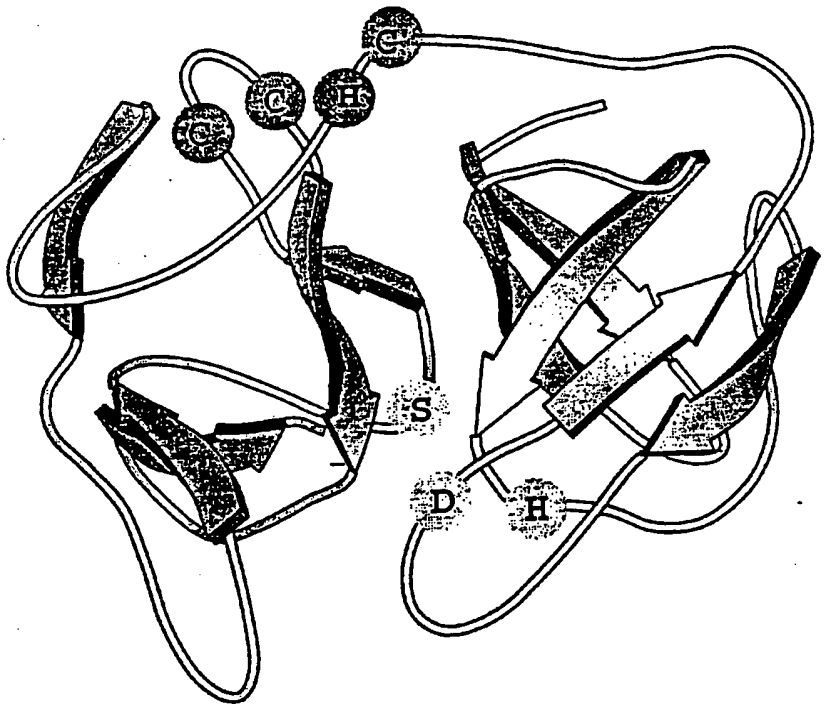

FIG. 2

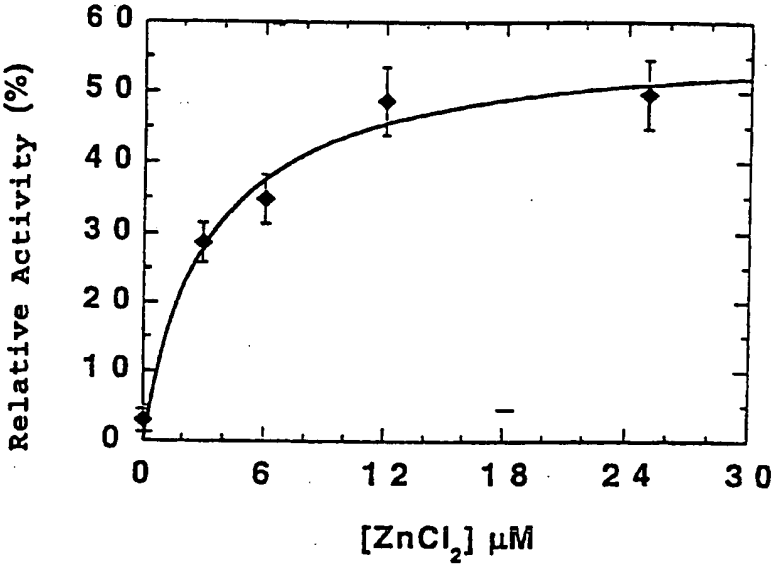

FIG. 3

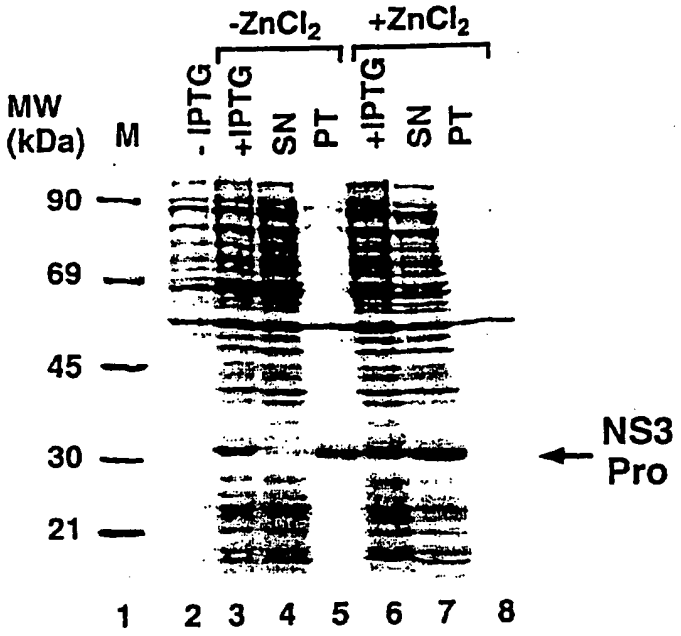
-ZnCl2
+ZnCl2
MW
(kDa)
M
- IPTG
+IPTG
SN
PT
+IPTG
SN
PT
90
69
45
30
21
NS3
Pro
1 2 3 4 5 6 7 8

FIG. 4

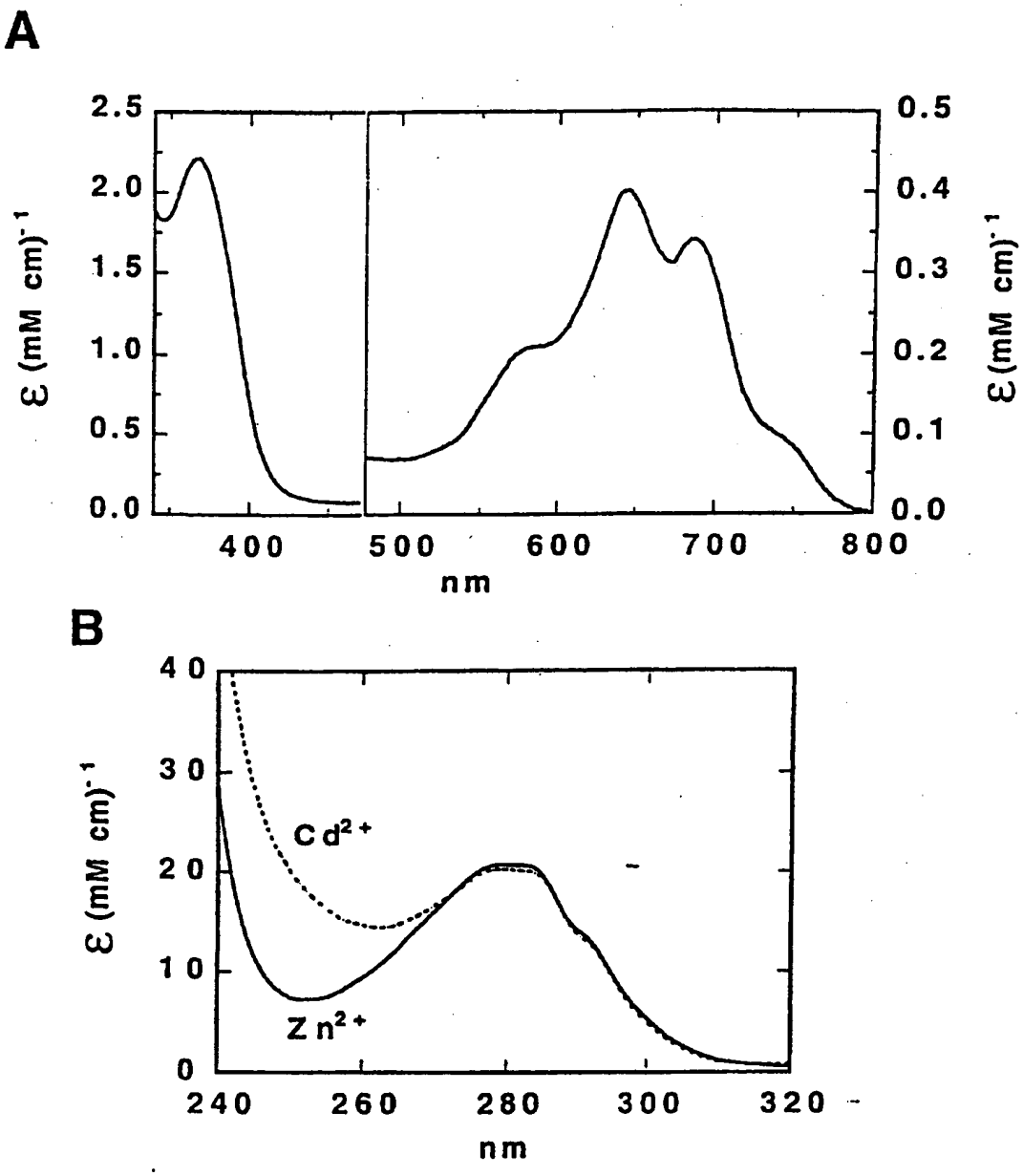
A
2.5
2.0
1.5
1.0
0.5
0.0
ε (mM cm)⁻¹
400
500
600
700
800
nm
0.5
0.4
0.3
0.2
0.1
0.0
ε (mM cm)⁻¹
B
40
30
20
10
0
ε (mM cm)⁻¹
Cd2+
Zn2+
240
260
280
300
320
nm

FIG. 5

INTERNATIONAL SEARCH REPORT

Intern. ..al Application No
PCT/IT 97/00228

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N9/50 C12N15/51 C12N15/62 C07K14/18

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C12N C07K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	KHUDYAKOV YU.E. ET AL: "Linear B-Cell Epitopes of the NS3-NS4-NS5 Proteins of the Hepatitis C Virus as modeled with Synthetic Peptides" VIROLOGY, vol. 206, 1995, pages 666-72, XP002003036 whole document, Table 1 NS3 first peptide ---	1-5
Y	WO 92 12992 A (GAMBLE JAMES N MED INST) 6 August 1992 whole document, esp. claims 25, 39e), 166, FGB30 domain page 13, line 20 --- -/--	1-5

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
14 January 1998	30. 01. 98

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Kronester-Frei, A

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Interna..ial Application No
PCT/IT 97/00228

C.(Continuation) DOCUMENTS CONSIDERED T BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	KIM D.W. ET AL.: "C-terminal Domain of Hepatitis C Virus NS3 protein contains an RNA helicas activity" BIOCHE.BIOPHYS.RESERARCH COMMUNICATIONS, vol. 215, no. 1, 1995, pages 160-166, XP002035618 see title	1-5
P,X	WO 97 08304 A (ANGELETTI P IST RICHERCHE BIO ;STEINKUEHLER CHRISTIAN (IT); PESSI) 6 March 1997 whole document, claims, Seq. ID 45	1-5
P,Y	WO 96 36702 A (SCHERING CORP) 21 November 1996 whole document, claims, Seq. ID 2-4 having C-terminal -RKKKRR	1-5
A	TOMEI L. ET AL.: "NS3 is a Serine Protease required for processing of Hepatitis C Virus Polyprotein" J. VIROLOGY, vol. 67, no. 7, 1993, pages 4017-4026, XP000561255 whole document	1-5

INTERNATIONAL SEARCH REPORT

Interna... ...al Application No
PCT/IT 97/00228

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9212992 A	06-08-92	AU 1459792 A	27-08-92
		EP 0571554 A	01-12-93
WO 9708304 A	06-03-97	IT RM950573 A	24-02-97
		AU 6668696 A	19-03-97
WO 9636702 A	21-11-96	AU 5729196 A	29-11-96